

Original citation:

Price, Jonathan, Harrison, M. C., Hammond, R. L., Adams, Stefan, Gutierrez-Marcos, José F. and Mallon, E. B. (2018) Alternative splicing associated with phenotypic plasticity in the bumble bee *Bombus terrestris*. *Molecular Ecology* . doi:[10.1111/mec.14495](https://doi.org/10.1111/mec.14495)

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/99883>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

"This is the peer reviewed version of the following Price, Jonathan, Harrison, M. C., Hammond, R. L., Adams, Stefan, Gutierrez-Marcos, José F. and Mallon, E. B. (2018) Alternative splicing associated with phenotypic plasticity in the bumble bee *Bombus terrestris*. *Molecular Ecology* . doi:[10.1111/mec.14495](https://doi.org/10.1111/mec.14495) which has been published in final form at <http://dx.doi.org/10.1111/mec.14495>. This article may be used for non-commercial purposes in accordance with [Wiley Terms and Conditions for Self-Archiving](#)."

A note on versions:

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP URL' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

Alternative splicing associated with phenotypic plasticity in the bumble bee *Bombus terrestris*

J. PRICE ¹, M.C. HARRISON ², R.L. HAMMOND ³, S. ADAMS ¹, J.F. GUTIERREZ-MARCOS ¹ & E.B. MALLON ^{3*}

1) School of Life Sciences
Gibbet Hill Campus
University of Warwick
Coventry, CV4 7AL
United Kingdom

2) Institute for Evolution and Biodiversity
University of Münster
Hüfferstraße 1
48149 Münster
Germany

3) Department of Genetics and Genome Biology
University of Leicester
University Road
Leicester, LE1 7RH
Phone: ++ (0)116 252 3488
Fax: +44 (0)116 252 3330
E-mail: ebm3@le.ac.uk

Keywords: worker, queen, isoform, eusociality, transcriptomics, reproductive division of labour

Running title: "Alternative splicing in the bumble bee."

Abstract

Phenotypic plasticity is when one genome can produce more than one phenotype. The caste system found in many social insects is an important example of plasticity. Several studies have examined gene expression in social insect developmental and caste differences. Changes in gene expression, however, are not the only source of phenotypic plasticity. Here we investigate the role of alternative splicing in the buff tailed bumble bee *Bombus terrestris*. We found that 5458 genes in *Bombus terrestris* (40%) express more than one isoform. Larvae have the lowest level of splicing events, followed by adults and then pupae. We found that when an isoform is expressed in a given caste in the larval stage, it tends to be expressed in all castes at the larval stage. The same is true at the pupal stage. However, we see more complicated interactions between the adult castes with reproductive females showing different isoform expression compared to non-reproductive females and male adults showing the most distinct patterns. We found 455 isoform switching genes, that is genes, where one developmental stage, sex or caste uses a specific isoform and another type uses a different isoform. Among genes displaying isoform switching are some involved in the ecdysteriod pathway, an important system in insect behaviour.

1 Introduction

2 A fundamental aim of evolutionary biology is to understand the link between genotype, phenotype and
3 fitness (de Visser and Krug, 2014). A powerful way to understand this is to investigate how a single
4 genome can give rise to multiple phenotypes. This question can be approached at the cellular level, for
5 example, how a single genome can make both a skin cell and a blood cell, but also at the societal level,
6 for example, how a single genome can make both the queen and worker castes in a eusocial insect colony.
7 Indeed, recent models suggest that these different levels could even be functionally similar (Patalano
8 *et al.*, 2012).

9 In eusocial insect colonies, this differentiation between castes is fundamental, with queens carrying
10 out most of the reproduction while workers carry out other tasks required of the colony e.g. brood
11 care, foraging and nest maintenance. Queens are also generally bigger, live longer and have a host of
12 behavioural and physiological specialisations compared to workers. These distinct morphological and
13 behavioural castes are usually alternative expressions of the same genome. We previously found that the
14 main difference in gene expression, in the buff-tailed bumble bee *Bombus terrestris*, was between castes at
15 the adult stage rather than at larval or pupal stages (Harrison *et al.*, 2015). A striking difference between
16 *B. terrestris* and higher eusocial hymenopterans is that reproductive worker bees have gene expression
17 profiles much more similar to queens than to their non-reproductive sister workers (Harrison *et al.*, 2015).

18 Changes in gene expression are not the only way a single genome can give rise to multiple phenotypes.
19 For example, the difference between head lice (harmless) and body lice, (disease vector) is caused by the
20 alternative splicing of an identical genome not by differential gene expression (Tovar-Corona *et al.*, 2015).
21 Alternative splicing results in a single gene coding for multiple processed messenger RNA (mRNA),
22 and potentially multiple proteins with distinct functions. Genome-wide analyses suggest alternative
23 splicing as a source of phenotypic variation in eusocial insects, where a single genome must encode for
24 numerous castes (honeybees (Foret *et al.*, 2012; Li-Byarlay *et al.*, 2013), possibly ants (Bonasio *et al.*,
25 2012) and termites (Terrapon *et al.*, 2014)). Bumble bees, as well as being one of the most ecologically and
26 economically important pollinators (Woodard *et al.*, 2015), are an interesting taxon to study the role of
27 alternative splicing in eusociality as they possess both highly eusocial and more primitive characteristics
28 (Harrison *et al.*, 2015).

29 The diversity of forms and behaviours within a eusocial insect species is much greater than that found
30 within solitary insect species. Yet eusocial insects have gene numbers within the range for solitary insects
31 (Simola *et al.*, 2013). Tovar-Corona *et al.* 2015 recently found the rate of alternative splicing in eight
32 arthropod species ranging from 23% to 44%. Do bumble bees use alternative splicing more than other
33 studied, non-social insects, in order to increase phenotypic diversity?

34 Here, using RNA-seq, we ask how alternative splicing is associated with caste, sex and development
35 within the buff-tailed bumble bee, *Bombus terrestris*. We compare splicing patterns of larval (male
36 versus workers), pupal (male versus workers) and adult (male versus non-reproductive worker versus
37 reproductive worker versus queen) stages. Using the results of this analysis we answer a number of
38 questions. Do bumble bees have increased rates of alternative splicing compared to studied non-social
39 insects? Do splicing patterns follow those of gene expression with relatively fewer differences between
40 juvenile stages compared to adults? Are reproductive workers' splicing patterns more similar to queens'
41 or non-reproductive workers'? Do different castes, sexes or developmental stages use different isoforms
42 of the same genes?

Methods

Colony sampling, RNA libraries and sequencing

Data for this study were obtained from Harrison *et al.* (2015). Reads were downloaded from the ENA under accession no. PRJEB9366. This contained a total of 469.3 million reads from 27 libraries. These libraries are made up of larval, pupal and adults stages of males, queens, workers (reproductive and non-reproductive). Whole bodies were used for sampling for two reasons. First, we had no prior assumptions regarding the tissues, within which genes would be alternatively spliced between castes, sexes and developmental stages. Second, this would allow us to detect as many alternatively spliced genes as possible across all comparisons. Most sample types are replicated over three colonies except for adult nonreproductive workers (two colonies), reproductive workers (four colonies) and larval, pupal and virgin queens (one colony). Sample production is described fully by Harrison *et al.* (2015). Due to the lack of replicates, queen larvae, pupae and virgin queen were not analysed in our current alternative splicing study.

Read processing and analysis

Raw reads were trimmed for adaptor sequence and low quality at the 5' and 3' end, and across a sliding window throughout the reads (PHRED >24) using Trimmomatic (v0.33) (Bolger *et al.*, 2014), then visualised using FastQC (v0.11.5). Reads were aligned to the Bter_1.0 genome (Refseq accession no. GCF_000214255.1 (Sadd *et al.*, 2015)) using Tophat (v2.1.0) (Trapnell *et al.*, 2009), with an average of 93.3% efficiency (82.8%-95.8%). After mapping, the protocol for the tuxedo package was followed as outlined in Trapnell *et al.* (2012). In brief, cufflinks (v2.2.1) was performed on the alignments from tophat, then cuffmerge (v1.0.0) was used to create a consensus transcriptome. Finally cuffdiff (v2.2.1) was used to quantify isoform expression and identify statistically significant changes in isoform expression between the samples. For this study we have analysed isoforms at the CD level because they represent a change in the protein translated. Further downstream analysis of these isoforms was then performed in R 3.2.3 (core Team, 2016) using the cummeRbund (v2.16.0) package and custom R scripts (<https://dx.doi.org/10.6084/m9.figshare.3201355>). rMATs (v3.2.5) was run with the alignments from tophat to identify the types of alternative splicing present (Shen *et al.*, 2014)

Isoform switching and Clustering

Relative isoform contributions for each gene were tested for differences pairwise between the samples utilising the jensen-shannon divergence test implemented by the cuffdiff program (FDR < 0.01). Allometric

differences in the amount of tissue in different castes, sexes or reproductive status could produce significant differences in isoform expression if judged solely on statistical testing. We therefore only consider isoforms with at least a log2-fold change of one in order to reduce the false-positive effects of tissue scaling (Montgomery and Mank, 2016). To identify those genes that display biologically significant isoform switching, the isoforms from significant genes identified by the jensen-shannon divergence test were subject to K-means clustering (Hartigan and Wong, 1979) (k=30-150 step of 10, Iterations = 500) based on their scaled relative abundance. The most distinct abundance patterns along with the fewest duplicated clusters were found at K of 110. Each cluster was then plotted separately to visualize isoform expression. As described by Aghamirzaie *et al.* 2015, clusters were then manually combined into super-clusters where they had similar relative abundance patterns.

Gene ontology (GO) analysis

We extracted the nucleotide sequence for all isoforms found in each one of the chosen super-clusters and searched for any matching sequence on NCBI using BLASTx (Altschul *et al.*, 1990) with an E-value cutoff of 0.001. Using Blast2Go (Gotz *et al.*, 2008), we identified gene ontology (GO) terms associated with these loci. We carried out an enrichment analysis (Fisher exact test) using a custom R script (<https://dx.doi.org/10.6084/m9.figshare.3201355.v1>) on each of these lists of super-cluster associated GO terms. This identified GO terms that are overrepresented ($FDR < 0.1$) relative to the entire transcriptome (<https://dx.doi.org/10.6084/m9.figshare.3458828.v1>).

Results

Isoform, CD and gene level quantification

We measured expression level in terms of fragments per kilobase of exon per million reads mapped (FPKM) of all isoforms present in published RNAseq libraries of several castes, sexes and developmental stages of *Bombus terrestris* (Figure 1). After removing genes with $\text{FPKM} < 1$ in all libraries there are 13548 distinct loci with expression in at least one sample. Alternative splicing produces different mRNA (isoforms) from the same gene. A subset of these isoforms will differ in their conserved protein domains (CD isoforms). At the isoform level, 8090 of the loci produce only one isoform, the remaining 5458 (40%) genes express 22786 isoforms ($\text{FPKM} > 0.5$ in at least one library), averaging 4.2 isoforms per gene. At the CD level 2833 (20.9%) show expression of more than one CD isoform in at least one sample, totalling 7726 CD isoforms (2.7 CD isoforms/gene). To avoid problems with assembly and annotation previously well described in alternative splicing studies (Hiller *et al.*, 2009; Jiang and Wong, 2009), we concentrated on the 2833 genes expressing more than one CD isoform (see Table 1).

CD isoform expression varies mainly between developmental stages, but also sex and caste

Expression (FPKM) for each of the 7726 CD isoforms is highly correlated between replicates of each biological type (mated queen, reproductive adult worker, non-reproductive adult worker, worker pupae, worker larvae, male adult, male pupae and male larvae) (Figure 1a) indicating that the patterns we observe here are of biological origin.

Principle component analysis of the 7726 CD isoform's expression (Figure 1b) shows that there is variance in the expression of these isoforms which is associated primarily with developmental stage but also sex and caste. PC1 principally describes variance in the isoform expression of different developmental stages by separating adult and pupal replicates on the left and larval replicates on the right, this difference explains 36% of the variance. PC2 is explained also by differences in developmental stages but this time separating the pupae samples from the rest (22% variance). PC3 displays the variance that exists between the sexes in the adult stages by dividing male adults and the female adult samples but also separating the female adult samples (13% variance).

Alternative splicing events can come in different forms; exon skipping, mutually exclusive exons, alternative 5' splice junction, alternative 3' splice junction and intron retention (Black, 2003). The proportions of the splicing event forms stay relatively constant between samples with only a small increase in skipped exon splice events for non-reproductive workers (see Figure S1). We observed differences in the numbers of spliced genes and the number of isoforms expressed in our samples (Table 1). Male and

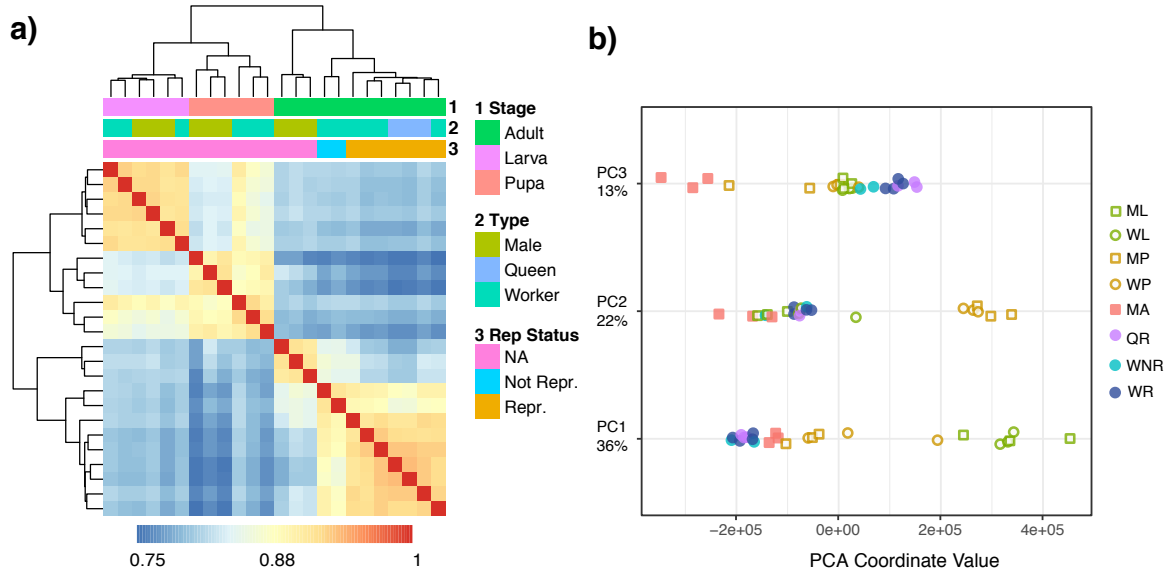


Figure 1: Global splicing analysis **a**. Correlogram of Pearson correlations of expression (FPKM) between replicates for the 7726 coding isoforms from multiple isoform loci. Clustering is produced based on Euclidean distances. This clustering separates replicates into stage, sex and reproductive status (coloured bars on top). **b**. A principal component analysis (PCA) of the 7726 isoforms from multiple isoform loci. QR = Queen Reproductive, WR = adult Worker Reproductive, WNR = adult Worker Non-Reproductive, WP = Worker Pupae, WL = Worker Larvae, MA = Male Adult, MP = Male Pupae, ML = Male Larvae.

Table 1: Total numbers of expressed features in the eight biological types examined

| Feature | QR | MA | ML | MP | WR | WNR | WL | WP |
|--|-------|-------|-------|-------|-------|-------|-------|-------|
| Genes | 9954 | 10043 | 9685 | 10756 | 10249 | 10418 | 9762 | 9969 |
| Genes With Multiple Isoforms | 5233 | 5295 | 5226 | 5338 | 5282 | 5311 | 5250 | 5265 |
| Isoforms from Genes with multiple isoforms | 16210 | 16840 | 16048 | 16844 | 16909 | 17090 | 16071 | 16274 |
| Genes with multiple CD isoforms | 2792 | 2801 | 2760 | 2800 | 2812 | 2818 | 2772 | 2782 |
| CD isoforms from genes with multiple CD isoforms | 6265 | 6326 | 6092 | 6349 | 6446 | 6472 | 6107 | 6208 |

worker larvae have the lowest number of transcripts expressed in each category (genes, multiple isoform genes, multiple isoforms, multiple CD genes, multiple CD isoforms). The pupae shows a discrepancy between the sexes with the male pupae expressing more transcripts of each type than the worker pupae. Male pupae express the most genes and the most multiple isoform genes but adult workers (WR and WNR) express the most isoforms, multiple CD isoform genes and the most CD isoforms.

455 genes with varied functions display significant isoform switching

The 2833 multiple isoform loci equate to the total amount of alternative splicing at the CD level observed in our samples but not all of these loci express different proportions of isoforms in different samples, that is, isoform switching. To find these loci, we tested the 2833 multiple isoform loci pairwise for differences

in their relative isoform abundance between samples ($\text{FDR} < 0.01$). 455 genes, 16% of the multiple isoform loci (2833) display significant isoform switching between samples. These 1303 CD isoforms from 455 genes were analysed further to identify biologically significant isoform switching events.

We used K-means clustering to identify clusters of isoforms displaying similar expression patterns across biological types. 297 (from 1303) were not clustered because they have small changes in their relative abundance between samples ($\text{range} < 0.2$) and would be misrepresented when their isoform abundance was scaled before clustering. K-means clustering ($K = 110$) of the 1006 remaining isoforms achieved separation of most isoforms by their relative abundance (Figure S2). By manually combining clusters with the same pattern of relative abundance (Figure 2) we identified 26 super-clusters that show a high degree of sample-specific expression (658 isoforms). The remaining 348 isoforms showed cryptic abundances and could not be assigned to a super-cluster, these isoforms are included in the variable cluster (Figure 2). 22 of the 26 super-clusters form antagonistic pairs, where both super-clusters represent the high and low abundance isoforms for that particular sample or group of samples. Most super-clusters (491 isoforms) show preferential high or low expression in samples(s) from one developmental stage and four clusters (167 isoforms) have preferential expression in samples from more than one developmental stage. The majority of isoforms show adult sample(s) specific expression (374) with 105 for pupae and 76 for larvae (supplementary data: isoform_cluster_database). Of the 455 isoform switching genes; 328 have multiple isoforms in multiple super-clusters, 39 have isoforms solely in the variable cluster, 25 genes have one or zero isoforms in a super-cluster and 63 are uncharacterised loci. From the 328 genes that have multiple isoforms in multiple super-clusters 80 genes have their isoforms clustered in antagonistic pairs.

For each super-cluster and groups of super-clusters relating to each developmental stage we tested for enriched gene ontology (GO) terms against their background value in the transcriptome. No discrete cluster had significant enriched GO terms at the $\text{FDR} < 0.1$. However, twenty three GO terms were associated with isoform switching genes found in adults (supplementary data: cluster_GO_terms). Two of these, steroid hormone mediated signalling pathway (GO:0043401) and steroid hormone receptor activity (GO:0003707) relate to important functions in social insect biology. A number of isoform switching genes are associated with these terms, including; ecdysone inducible protein 74EF (LOC100651988), E75 (LOC100644185), E78 (LOC100646780), ecdysone receptor (LOC100646757) and farnesyl pyrophosphatase synthase-like (LOC00165005). The majority of these genes differ in their isoform composition between one or more adult sample. As an example of one of these genes, we show the expression, exon map and protein domains for ecdysone inducible protein 75 (Figure 3). In this gene, transcript variant X1 (XM_003398944.2) has high levels of expression especially for non-reproductive females compared to transcript variant X2 (XM_003398942.2) (see Figure 3b). Expression plots and exon maps

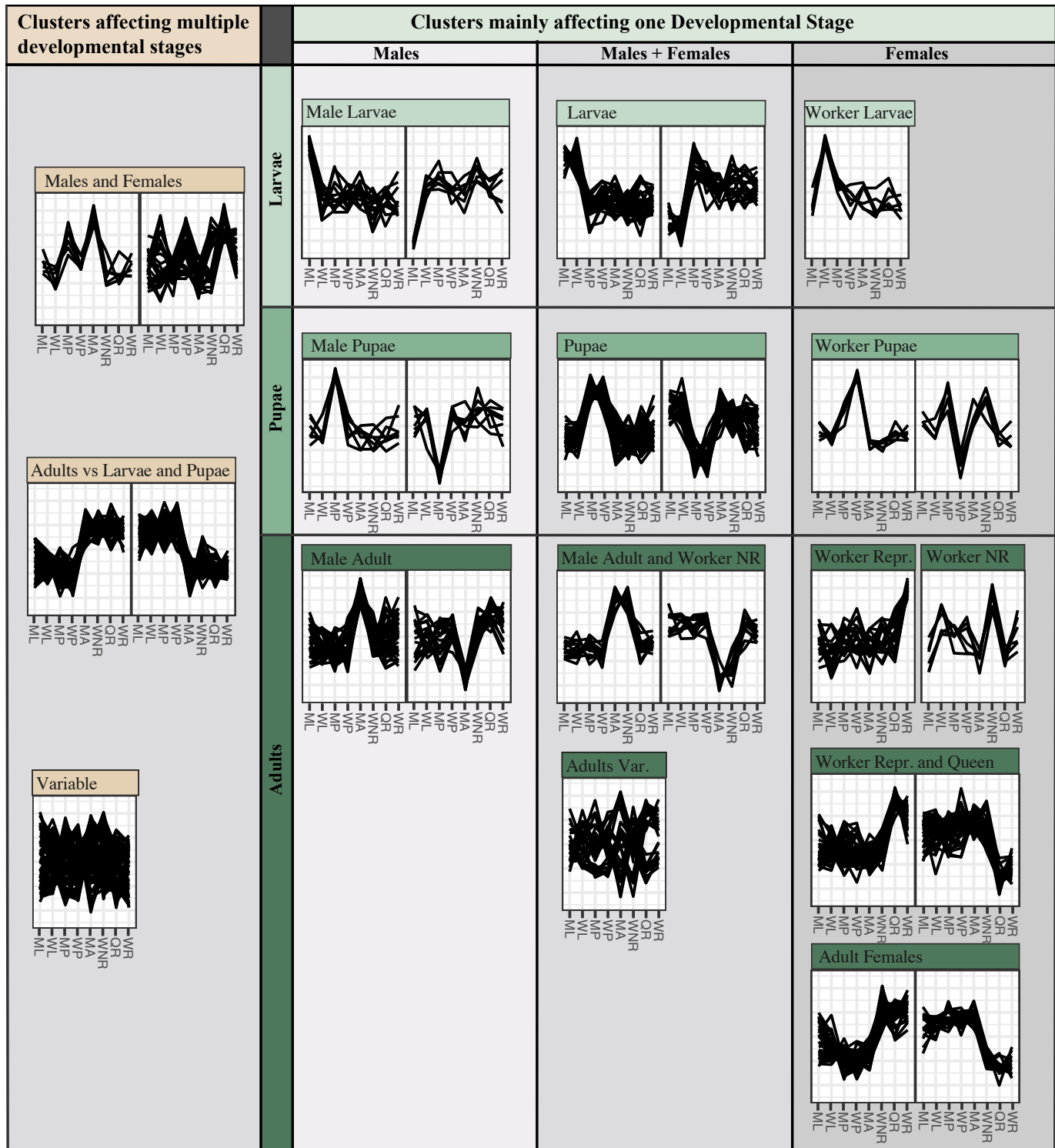


Figure 2: Super-clusters showing abundance patterns around particular biological sample types. K-means clustering for the 1006 isoforms from differentially spliced genes. Y-axis displays scaled relative abundance (-2 to 2). QR = Queen Reproductive, WR = adult Worker Reproductive, WNR = adult Worker Non-Reproductive, WP = Worker Pupae, WL = Worker Larvae, MA = Male Adult, MP = Male Pupae, ML = Male Larvae. The labels in the banners above figures identify the biological types in which the cluster is differentially expressed in. For example, Male Pupae (second column, second row) shows two clusters, one where there is a spike in expression in male pupae and one where there is a drop.

165 (10.6084/m9.figshare.5208940) are available in the supplementary data for all 455 isoform switching genes
166 along with a database containing their annotation (supplementary data: isoform_cluster_database),
167 their isoform's super-cluster membership (supplementary data: Clusters in super clusters) and GO terms
168 (supplementary data: cluster_GO_terms).

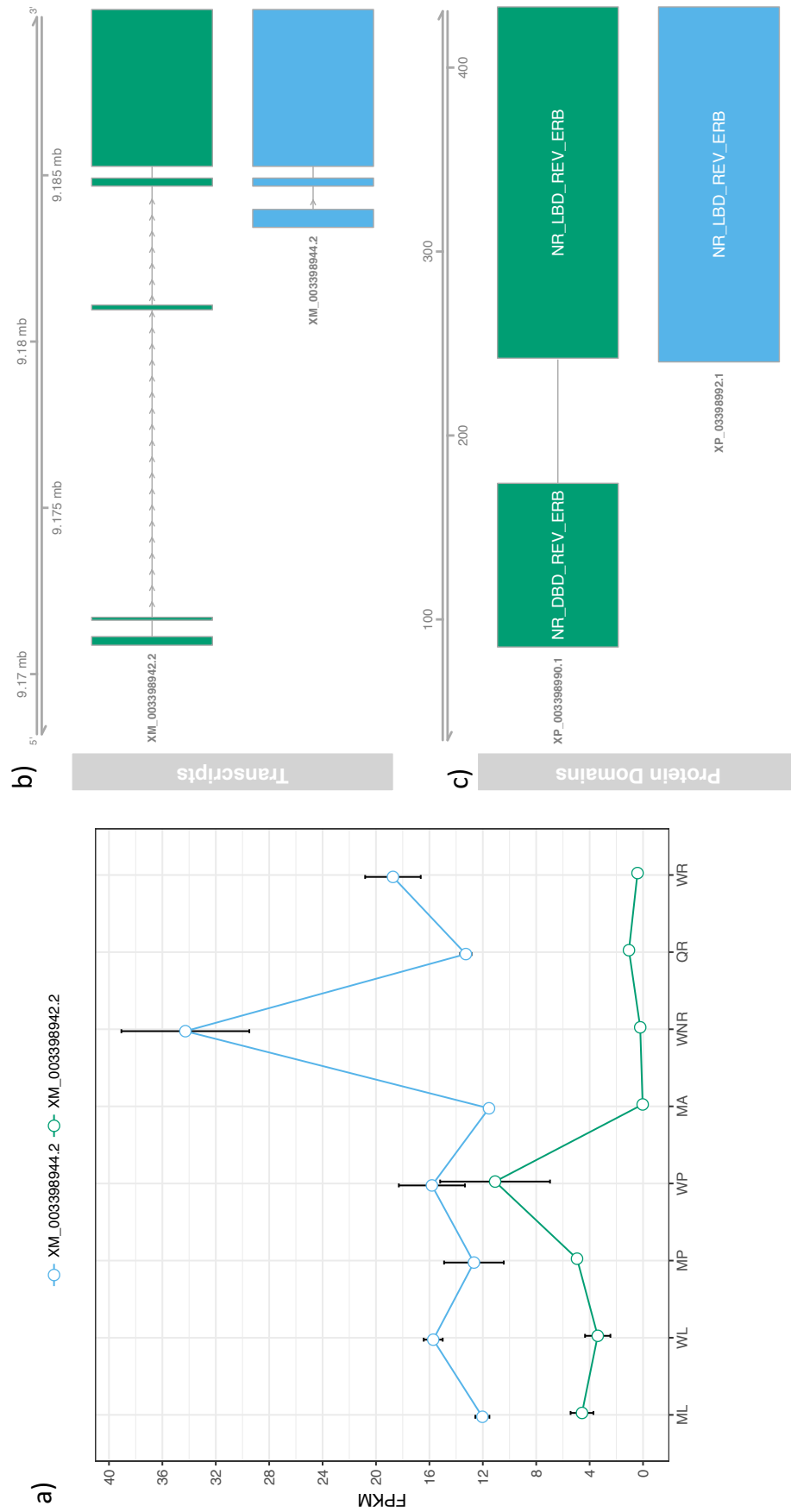


Figure 3: **a.** Expression (FPKM) of the two transcripts of *Ecdysone-inducible protein E75* over the different castes and developmental stages. Circles represent the means, the error bars are standard errors of the mean. QR = Queen Reproductive, WR = adult Worker Reproductive, WNR = adult Worker Non-Reproductive, WP = Worker Pupae, WL = Male Adult, MP = Male Pupae, ML = Male Larvae **b** Exon map of the two transcripts. **c** Protein domains encoded by the two transcripts. NR_LBD_REV_ERB = ligand binding domain of REV-ERB receptors, NR_DBD_REV_ERB = DNA-binding domain of REV-ERB receptor-like.

Discussion

Alternative splicing is a way to increase the variety of proteins produced by a given number of genes. One aim of this paper was to ask do bumble bees have increased rates of alternative splicing to account for the increased phenotypic diversity within a eusocial insect species despite them having gene numbers within the range for solitary insects (Simola *et al.*, 2013). We observed 2833 genes that have expression of more than one CD isoform in the samples examined. This corresponds to 20% (2833/13548) of the bumble bee genes producing different proteins from the same genotype. To compare *Bombus terrestris* with current literature we also looked at the number of genes that express more than one isoform regardless of the protein produced (Tovar-Corona *et al.*, 2015). Compared to the rate of alternative splicing in eight arthropod species recently studied (23% to 44%) (Tovar-Corona *et al.*, 2015), bumble bees' rate of 40% is at the higher end. However this rate is equivalent to those of two other insects (*Drosophila melanogaster* 42% and the pea aphid, *Acyrtosiphon pisum* 44%). Further work in other insects will be needed to confirm this, but our data indicates that an increased rate of alternative splicing is not required by bumble bees.

We then examined how the bumble bee uses these isoforms. We found four hundred and fifty five genes express different proportions of isoforms in different samples, that is, isoform switching. We have annotated and ranked these genes according to their super cluster membership, conserved domain changes in their isoforms, GO terms and gene expression status, providing a rich resource for future studies in this and related species. Patterns of alternative splicing follow those previously found for gene expression (Harrison *et al.*, 2015). This is not due to the effect of overall expression itself, as we only looked at genes that have expression of at least two isoforms, removing almost 80% of the observed genes (13548 - 2,833). Even still, within the 455 isoform switching genes, 45% also show gene expression changes. This exemplifies how alternative splicing studies can both add to the knowledge of genes known to be differentially expressed and also find novel source of plasticity that could not be found by only examining gene level expression.

We found relatively little differentiation between the sexes at the larval stage when compared to adult stages (see Figures 1 and 2). These findings suggest a comparatively low number of isoforms are associated with the creation of distinct morphological sexes compared to the high number involved in distinct behaviours between adult males and females. Reproductive workers show more similarity to queens than to their fellow, but non-reproductive, workers (see horizontal dendrograms Figure 1a). We have hypothesised that this is due to the plastic nature of bumble bee workers which allow them to compete with the queen at the end of the annual colony cycle (Harrison *et al.*, 2015). Male adults are the most distinct group in terms of isoform expression (Figure 1), a finding that reflects their similarly

distinct behaviour, physiology and life history.

No GO terms were found to be significantly associated with any of the super clusters. However, twenty three GO terms were associated with isoform switching genes found in adults. Two of these, steroid hormone mediated signalling pathway (GO:0043401) and steroid hormone receptor activity (GO:0003707) relate to the ecdysteroid pathway, an important system in insect biology (Niwa and Niwa, 2016). Ecdysteroids are key compounds involved in ovary activation, regulating agonistic behaviour and establishing the dominance hierarchy in bumble bee workers and queens (Geva *et al.*, 2005).

In the results, as an example of an isoform switching gene, we expanded our analysis for one ecdysone related gene (Ecdysone-inducible protein E75 (LOC100644185)). Ecdysone-inducible protein E75 is a transcription factor in the ecdysone cascade (Li *et al.*, 2015). Transcript variant X1 (XM_003398944.2) has high levels of expression especially for non-reproductive females compared to transcript variant X2 (XM_003398942.2) (see Figure 3b). In *Drosophila melanogaster*, different isoforms of E75 have opposite roles in controlling female reproduction with, for example, DmE75A inducing apoptosis in the egg chamber, while DmE75B inhibits DmE75A function allowing egg development (Terashima and Bownes, 2006). The difference between the transcripts is a number of exons encoding for binding domains of the nuclear receptor REV-ERB (Figure 3c). REV-ERB receptors are transcriptional regulators belonging to the nuclear receptor superfamily. They regulate a number of physiological functions including the circadian rhythm, lipid metabolism, and cellular differentiation. The ligand binding domain (LBD) of REV-ERB receptor-like, found in both isoforms, is responsible for coactivator interaction with steroid hormones (Raghuram *et al.*, 2007). The DNA-binding domain (DBD), found only in transcript variant X2, interacts with specific DNA sites upstream of the target gene and modulates the rate of transcriptional initiation. This truncation of a binding domain is a very simple way for a gene to go from being expressed specifically to a more global pattern of expression. We found this truncation in two other genes we examined in depth, the CXXC zinc finger domain in Methylcytosine dioxygenase TET2 (LOC100642293 Figure S3) and the RNA recognition motif in Insulin-like growth factor 2 mRNA-binding protein 1 (LOC100647990 Figure S4). Further study of our data set and others could test if this is a common phenomenon.

We have shown that alternative splicing of genes is important for bumble bee biology. How is this alternative splicing controlled? Alternative splicing in some social insect species has been shown to be affected by DNA methylation (Bonasio *et al.*, 2012; Foret *et al.*, 2012). Methylation has also been shown to play a role in social insect biology (Yan *et al.*, 2015) including bumble bee reproductive division of labour (Amarasinghe *et al.*, 2014). Together these observations suggest that a fruitful area of research for social insect biologists would be in understanding the role of methylation in controlling alternative splicing.

Acknowledgements

JP was supported by a a PhD research grant from the BBSRC. MCH was supported by a PhD research grant from the Natural Environment Research Council (NERC). EBM was supported by research grant NE/N010019/1 from the Natural Environment Research Council, UK. The raw data used in this study was funded by a NERC Biomolecular Analysis Facility research grant (NBAF 829). Illumina library preparation and sequencing were carried out by Edinburgh Genomics, The University of Edinburgh. Edinburgh Genomics is partly supported through core grants from NERC (R8/H10/56), MRC (MR/K001744/1) and BBSRC (BB/J004243/1).

References

- Aghamirzaie, D., Batra, D., Heath, L. S., Schneider, A., Grene, R., and Collakova, E. 2015. Transcriptome-wide functional characterization reveals novel relationships among differentially expressed transcripts in developing soybean embryos. *BMC Genomics*, 16.
- Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D. 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3): 403–410.
- Amarasinghe, H. E., Clayton, C. I., and Mallon, E. B. 2014. Methylation and worker reproduction in the bumble-bee (*Bombus terrestris*). *Proceedings of the Royal Society B: Biological Sciences*, 281(1780): 20132502.
- Black, D. L. 2003. Mechanisms of Alternative Pre-Messenger RNA Splicing.
- Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15): 2114–2120.
- Bonasio, R., Li, Q., Lian, J., Mutti, N. S., Jin, L., Zhao, H., Zhang, P., Wen, P., Xiang, H., Ding, Y., Jin, Z., Shen, S. S., Wang, Z., Wang, W., Wang, J., Berger, S. L., Liebig, J., Zhang, G., and Reinberg, D. 2012. Genome-wide and Caste-Specific DNA Methylomes of the Ants *Camponotus floridanus* and *Harpegnathos saltator*. *Current Biology*, 22(19): 1755–1764.
- core Team, R. 2016. R: A language and environment for statistical computing.
- de Visser, J. A. G. M. and Krug, J. 2014. Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics*, 15(7): 480–490.
- Foret, S., Kucharski, R., Pellegrini, M., Feng, S., Jacobsen, S. E., Robinson, G. E., and Maleszka, R. 2012. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *Proceedings of the National Academy of Sciences*, 109(13): 4968–4973.
- Geva, S., Hartfelder, K., and Bloch, G. 2005. Reproductive division of labor, dominance, and ecdysteroid levels in hemolymph and ovary of the bumble bee *Bombus terrestris*. *Journal of Insect Physiology*, 51(7): 811–823.
- Gotz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J., and Conesa, A. 2008. High-throughput functional annotation and data mining with the Blast2go suite. *Nucleic Acids Research*, 36(10): 3420–3435.
- Harrison, M. C., Hammond, R. L., and Mallon, E. B. 2015. Reproductive workers show queen-like gene expression in an intermediately eusocial insect, the buff-tailed bumble bee *Bombus terrestris*. *Molecular Ecology*, 24: 121–129.
- Hartigan, J. A. and Wong, M. A. 1979. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1): 100–108.
- Hiller, D., Jiang, H., Xu, W., and Wong, W. H. 2009. Identifiability of isoform deconvolution from junction arrays and RNA-Seq. *Bioinformatics (Oxford, England)*, 25(23): 3056–3059.
- Jiang, H. and Wong, W. H. 2009. Statistical inferences for isoform expression in RNA-Seq. *Bioinformatics (Oxford, England)*, 25(8): 1026–1032.
- Li, K., Guo, E., Hossain, M. S., Li, Q., Cao, Y., Tian, L., Deng, X., and Li, S. 2015. Bombyx E75 isoforms display stage- and tissue-specific responses to 20-hydroxyecdysone. *Scientific Reports*, 5: 12114.
- Li-Byarlay, H., Li, Y., Stroud, H., Feng, S., Newman, T. C., Kaneda, M., Hou, K. K., Worley, K. C., Elsik, C. G., Wickline, S. A., Jacobsen, S. E., Ma, J., and Robinson, G. E. 2013. RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee. *Proceedings of the National Academy of Sciences*, 110(31): 12750–12755.
- Montgomery, S. H. and Mank, J. E. 2016. Inferring regulatory change from gene expression: the confounding effects of tissue scaling. *Molecular Ecology*, 25(20): 5114–5128.

- 289 Niwa, Y. S. and Niwa, R. 2016. Transcriptional regulation of insect steroid hormone biosynthesis and
290 its role in controlling timing of molting and metamorphosis. *Development, Growth & Differentiation*,
291 58(1): 94–105.
- 292 Patalano, S., Hore, T., Reik, W., and Sumner, S. 2012. Shifting behaviour: Epigenetic reprogramming
293 in eusocial insects. *Current Opinion in Cell Biology*, 24(3): 367–373.
- 294 Raghuram, S., Stayrook, K. R., Huang, P., Rogers, P. M., Nosie, A. K., McClure, D. B., Burris, L. L.,
295 Khorasanizadeh, S., Burris, T. P., and Rastinejad, F. 2007. Identification of heme as the ligand for the
296 orphan nuclear receptors REV-ERBalpha and REV-ERBbeta. *Nature Structural & Molecular Biology*,
297 14(12): 1207–1213.
- 298 Sadd, B. M., Barribeau, S. M., Bloch, G., Graaf, D. C. d., Dearden, P., Elsik, C. G., Gadau, J., Grim-
299 melikhuijzen, C. J., Hasselmann, M., Lozier, J. D., Robertson, H. M., Smagghe, G., Stolle, E., Vaeren-
300 bergh, M. V., Waterhouse, R. M., Bornberg-Bauer, E., Klasberg, S., Bennett, A. K., Câmara, F.,
301 Guigó, R., Hoff, K., Mariotti, M., Munoz-Torres, M., Murphy, T., Santesmasses, D., Amdam, G. V.,
302 Beckers, M., Beye, M., Biewer, M., Bitondi, M. M., Blaxter, M. L., Bourke, A. F., Brown, M. J.,
303 Buechel, S. D., Cameron, R., Cappelle, K., Carolan, J. C., Christiaens, O., Ciborowski, K. L., Clarke,
304 D. F., Colgan, T. J., Collins, D. H., Cridge, A. G., Dalmay, T., Dreier, S., Plessis, L. d., Duncan, E.,
305 Erler, S., Evans, J., Falcon, T., Flores, K., Freitas, F. C., Fuchikawa, T., Gempe, T., Hartfelder, K.,
306 Hauser, F., Helbing, S., Humann, F. C., Irvine, F., Jermini, L. S., Johnson, C. E., Johnson, R. M.,
307 Jones, A. K., Kadowaki, T., Kidner, J. H., Koch, V., Köhler, A., Kraus, F. B., Lattorff, H. M., Leask,
308 M., Lockett, G. A., Mallon, E. B., Antonio, D. S. M., Marxer, M., Meeus, I., Moritz, R. F., Nair, A.,
309 Näpflin, K., Nissen, I., Niu, J., Nunes, F. M., Oakeshott, J. G., Osborne, A., Otte, M., Pinheiro, D. G.,
310 Rossié, N., Rueppell, O., Santos, C. G., Schmid-Hempel, R., Schmitt, B. D., Schulte, C., Simões, Z. L.,
311 Soares, M. P., Swevers, L., Winnebeck, E. C., Wolschin, F., Yu, N., Zdobnov, E. M., Aqrabi, P. K.,
312 Blankenburg, K. P., Coyle, M., Francisco, L., Hernandez, A. G., Holder, M., Hudson, M. E., Jackson,
313 L., Jayaseelan, J., Joshi, V., Kovar, C., Lee, S. L., Mata, R., Mathew, T., Newsham, I. F., Ngo, R.,
314 Okwuonu, G., Pham, C., Pu, L.-L., Saada, N., Santibanez, J., Simmons, D., Thornton, R., Venkat,
315 A., Walden, K. K., Wu, Y.-Q., Debyser, G., Devreese, B., Asher, C., Blommaert, J., Chipman, A. D.,
316 Chittka, L., Fouks, B., Liu, J., O'Neill, M. P., Sumner, S., Puiu, D., Qu, J., Salzberg, S. L., Scherer,
317 S. E., Muzny, D. M., Richards, S., Robinson, G. E., Gibbs, R. A., Schmid-Hempel, P., and Worley,
318 K. C. 2015. The genomes of two key bumblebee species with primitive eusocial organization. *Genome*
319 *Biology*, 16(1): 76.
- 320 Shen, S., Park, J. W., Lu, Z.-x., Lin, L., Henry, M. D., Wu, Y. N., Zhou, Q., and Xing, Y. 2014.
321 rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data.
322 *Proceedings of the National Academy of Sciences*, 111(51): E5593–E5601.
- 323 Simola, D. F., Wissler, L., Donahue, G., Waterhouse, R. M., Helmkampf, M., Roux, J., Nygaard, S.,
324 Glastad, K., Hagen, D. E., Viljakainen, L., Reese, J. T., Hunt, B. G., Graur, D., Elhaik, E., Kriventseva,
325 E., Wen, J., Parker, B. J., Cash, E., Privman, E., Childers, C. P., Munos-Torres, M. C., Boomsma, J. J.,
326 Bornberg-Bauer, E., Currie, C., Elsik, C. G., Suen, G., Goodisman, M. A., Keller, L., Liebig, J., Rawls,
327 A., Reinberg, D., Smith, C. D., Smith, C. R., Tsutsui, N., Wurm, Y., Zdobnov, E. M., Berger, S. L., and
328 Gadau, J. 2013. Social insect genomes exhibit dramatic evolution in gene composition and regulation
329 while preserving regulatory features linked to sociality. *Genome Research*, page gr.155408.113.
- 330 Terashima, J. and Bownes, M. 2006. E75a and E75b have opposite effects on the apoptosis/development
331 choice of the *Drosophila* egg chamber. *Cell Death and Differentiation*, 13(3): 454–464.
- 332 Terrapon, N., Li, C., Robertson, H. M., Ji, L., Meng, X., Booth, W., Chen, Z., Childers, C. P., Glastad,
333 K. M., Gokhale, K., Gowin, J., Gronenberg, W., Hermansen, R. A., Hu, H., Hunt, B. G., Huylmans,
334 A. K., Khalil, S. M. S., Mitchell, R. D., Munoz-Torres, M. C., Mustard, J. A., Pan, H., Reese, J. T.,
335 Scharf, M. E., Sun, F., Vogel, H., Xiao, J., Yang, W., Yang, Z., Yang, Z., Zhou, J., Zhu, J., Brent,
336 C. S., Elsik, C. G., Goodisman, M. A. D., Liberles, D. A., Roe, R. M., Vargo, E. L., Vilcinskas, A.,
337 Wang, J., Bornberg-Bauer, E., Korb, J., Zhang, G., and Liebig, J. 2014. Molecular traces of alternative
338 social organization in a termite genome. *Nature Communications*, 5: 3636.

- 339 Tovar-Corona, J. M., Castillo-Morales, A., Chen, L., Olds, B. P., Clark, J. M., Reynolds, S. E., Pitten-
340 drigh, B. R., Feil, E. J., and Urrutia, A. O. 2015. Alternative Splice in Alternative Lice. *Molecular*
341 *Biology and Evolution*, 32(10): 2749–2759.
- 342 Trapnell, C., Pachter, L., and Salzberg, S. L. 2009. TopHat: discovering splice junctions with RNA-Seq.
343 *Bioinformatics*, 25(9): 1105–1111.
- 344 Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L.,
345 Rinn, J. L., and Pachter, L. 2012. Differential gene and transcript expression analysis of RNA-seq
346 experiments with TopHat and Cufflinks. *Nature Protocols*, 7(3): 562–578.
- 347 Woodard, S. H., Lozier, J. D., Goulson, D., Williams, P. H., Strange, J. P., and Jha, S. 2015. Molecular
348 tools and bumble bees: revealing hidden details of ecology and evolution in a model system. *Molecular*
349 *Ecology*, 24(12): 2916–36.
- 350 Yan, H., Bonasio, R., Simola, D. F., Liebig, J., Berger, S. L., and Reinberg, D. 2015. DNA Methylation in
351 Social Insects: How Epigenetics Can Control Behavior and Longevity. *Annual Review of Entomology*,
352 60(1): 435–452.

353 Data Accessibility,

354 All sequence data for this study are archived at European Nucleotide Archive (ENA); Accession no. PR-
355 JEB9366 (<http://www.ebi.ac.uk/ena/data/view/PRJEB9366>). GO-analysis results and lists of isoforms
356 are available as Supporting Information.

357 Author contributions

358 All authors developed the project idea and were involved in the interpretation of data and finalization of
359 the manuscript. JP and EBM analyzed the data. EBM and JP drafted the manuscript.